Molecular chaperones: **Avoiding the crowd**

R. John Ellis

The involvement of two types of molecular chaperone in folding newly synthesized proteins can be rationalized in terms of the crowded nature of the intracellular environment. Recent work sheds light on how these chaperones recognise their substrates and protect them from the problems of macromolecular crowding.

Address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.
Email: je@dna.bio.warwick.ac.uk

Current Biology 1997, 7:R531–R533
http://biomednet.com/elecref/09609822007R0531

One of the criticisms that can be made of biochemistry is the tendency of some practitioners to assume too readily that interactions observed in their test tubes are directly relevant to what happens inside the living cell. A major difference between cells and test tubes is that the intracellular environment is highly crowded because of the high concentration of macromolecules — of the order of 200–400 mg ml^{-1} of protein and RNA — that it contains. Biochemical studies, by contrast, are commonly made using highly dilute solutions. This difference matters, because one effect of crowding is to increase the thermodynamic activities of macromolecules by several orders of magnitude, prompting the suggestion that "biochemical rates and equilibria in a living organism may bear scant resemblance to those measured in a bath of solvent" [1].

This effect on thermodynamic activity results simply from the fact that the concentration of macromolecules inside cells is so high that a significant fraction of the volume is physically occupied by macromolecules, and so is unavailable to other macromolecules. Theory predicts that the thermodynamic activity of each macromolecular species in a crowded environment exceeds the activity of the same species in dilute solution, while experimentation shows that the ratio of thermodynamic activity to concentration — the activity coefficient — for a spherical protein of 30 nm radius inside *Escherichia coli* is in the range 100–1000, depending on the effective specific volume assumed for the background macromolecules [2]. Thus the association constant for two proteins that are interacting inside the cell is greater than that measured for the same proteins interacting in the test tube.

There exists a well developed biophysical theory of macromolecular crowding, but to date its impact on biochemistry has been largely confined to studies of DNA replication [1]. Recent reports show the importance of considering this crowding phenomenon when trying to understand why the folding of some polypeptides inside cells involves two distinct types of molecular chaperone (reviewed in [3–5]).

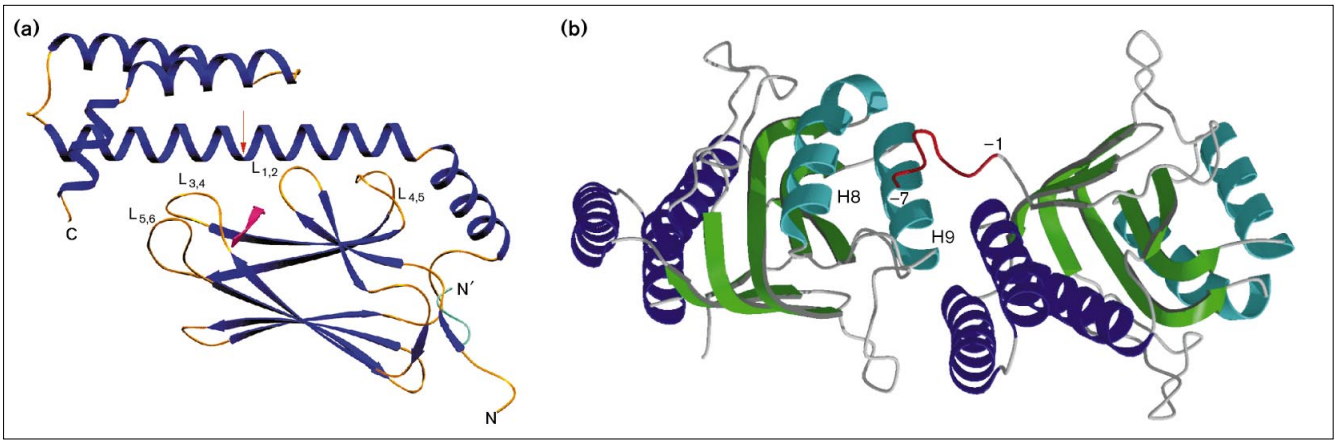
Two types of chaperone

These two types of molecular chaperone can for brevity be called small and large [5]; each type has a distinct function [3,4]. Small chaperones of the stress protein DnaJ and DnaK (Hsp70) families bind as monomers of 40 kDa and 70 kDa, respectively, to short extended runs of hydrophobic residues as they appear on elongating nascent polypeptide chains.This binding is transient, being regulated by slow ATP hydrolysis by the DnaK component, and serves both as a delaying device to prevent premature folding before a complete folding unit has been synthesized, and as an anti-aggregation device to prevent interaction with hydrophobic residues on adjacent nascent chains.

Release of these small chaperones leaves the polypeptide chain in an unfolded state until a sufficient length of chain to fold into a domain has been synthesized; consecutive domains may fold before the chain is complete. If these folded domains expose insufficient surface hydrophobic residues to cause aggregation, the chain can be released safely into the cytoplasm. In some cases, however, a domain — especially one that folds slowly — adopts a partially folded compact intermediate state, or 'molten globule', that exposes sufficient surface hydrophobic residues to run the risk of aggregating with either adjacent domains or domains in similar compact states released from the same or nearby polysomes. This problem is likely to affect only a subset of chains that complete their folding post-translationally.

The risk of aggregation is greatly increased by the effect of macromolecular crowding, as aggregation is a high order reaction. The concentration of ribosomes inside *E. coli* is estimated to be about 30 μM; as at any one time most ribosomes are active in protein synthesis, the concentration of nascent chains will also be about 30 μM, or almost 1 mg ml^{-1} for a 30 kDa chain. But the effect of crowding is predicted to increase the effective concentration to 3–30 mM or 100–1000 mg ml^{-1}! Protein chemists studying the refolding of pure denatured proteins in the test tube commonly reduce the protein concentration to below 1 mg ml^{-1} to avoid the problem of aggregation. The cell cannot do this, nor can it use the small chaperones for this purpose, as they bind only to extended hydrophobic residues. So another type of chaperone that can recognise hydrophobic residues on the surface of molten globules is required to act as an anti-aggregation device specific for such partially folded intermediates.

Figure 1



Comparison of the peptide binding site in the small chaperone DnaK with that in the large chaperone GroEL. **(a)** A ribbon diagram of DnaK peptide binding domain, with peptide (pink) bound in a channel formed by loops – $L_{1,2}$ and $L_{3,4}$ – of the β sandwich and capped by an α-helical domain. The red arrow indicates the point at which the long αB helix kinks up in one of the two known crystal forms. The amino-terminal residues can take up either of two alternative conformations, the second (N′) being shown in blue-green. The opening of the peptide binding cleft in the β sheet subdomain is seen from the side, with the long axis of the cleft perpendicular to the plane of the page. (Image generously provided and prepared by Wayne Hendrickson and Xiaotian Zhu.) **(b)** Ribbon diagram of the isolated apical domain of GroEL showing the amino-terminal heptapeptide tag (labelled −1 to −7 and coloured red) interacting with the neighbouring molecule in the crystal lattice. (Graphic kindly provided by Ashley Buckle; adapted from [12].)

Genetic evidence [6] supports the view that large chaperones of the chaperonin (GroEL/ES) family [7] act to prevent protein aggregation inside an *E. coli* cell. Each molecule of the oligomeric chaperonin 60 component — GroEL, of mass about 805 kDa — contains two large central cavities, one at each end of the barrel-shaped molecule. Each GroEL cavity is lined by the apical domains of seven ATPase subunits, which bind to hydrophobic patches on the surface of one molten globule. This non-covalent interaction prevents the bound molten globule from aggregating with other molten globules. *In vitro* experiments suggest that a side effect of this anti-aggregation role is the partial unfolding of non-aggregated chains that have become kinetically trapped in misfolded conformations, thus allowing such chains another chance to fold correctly [8]. Unlike its anti-aggregation role, however, there is no evidence as yet that the unfolding effect of the chaperonin is important *in vivo*.

Having prevented aggregation, the chaperonin must now contrive to allow the molten globule to fold further to the point where it can be safely released into the cytoplasm. This is achieved by the binding to one end of the chaperonin 60 of the other large chaperone, chaperonin 10 — GroES, of mass about 98 kDa. This binding results in the the release of the molten globule into the central cavity of chaperonin 60 that is capped by chaperonin 10. This capping physically prevents the released globule from diffusing into the cytoplasm, which it would otherwise do at a rate much faster than the rate of folding [3,4,9].

Inside this enclosed cavity, which I have termed an Anfinsen cage [9], folding proceeds for a time set by an ATPase timer that releases the chaperonin 10 cap. This time is about 10 seconds at 37°C, or about half the time it takes to synthesize the average polypeptide inside *E. coli*. This makes sense as, in a rapidly growing cell, protein folding must keep pace with protein synthesis if partially folded intermediates prone to aggregation are not to accumulate. The protein is then free to either diffuse into the cytoplasm or to rebind to the apical domains if sufficient hydrophobic residues are still exposed, so that another round of folding is necessary.

Uncertainties

Two questions about the above view of protein folding in the cell have been under recent scrutiny. Firstly, what is the structural basis for the difference in specificity between the small and the large chaperones? Secondly, what is the biological significance of the *in vitro* observation that each oligomer of chaperonin 60 binds only one oligomer of chaperonin 10, so that molten globules bound to cavities not capped by chaperonin 10 are released into the free solution before rebinding to another molecule of chaperonin 60 [10]? Recent papers address both these problems.

The crystal structure of the two domain peptide-binding unit of DnaK from *E. coli* has been determined to 2 Å resolution [11]. This unit contains a synthetic heptapeptide (sequence NRLLLTG) bound in extended conformation through a channel defined by loops extended from

a β sandwich subdomain (Figure 1a). Binding interactions centre on a central leucine that is completely buried in a deep pocket. The peptide interacts with the DnaK peptide-binding unit through many van der Waals contacts made by its side chains, as well as through seven main-chain hydrogen bonds and additional main-chain van der Waals contacts.

Adjacent to the peptide-binding channel in DnaK is a set of α helices that is interpreted to act as a flexible lid to encapsulate the bound peptide. This interpretation is based on conformational differences seen in this subdomain in two different crystal forms, and can be rationalized in terms of the need to prevent the bound hydrophobic residues from interacting with similar residues in adjacent peptides. In this respect, peptide binding to DnaK is different from the binding of peptides to major histocompatibility complex (MHC) molecules, where the peptide lies in a open groove so as to be available for antigen presentation. Zhu *et al.* [11] suggest that the helical lid domain is closed when ADP is bound, but opens to allow peptide release in the ATP-bound state.

The crystal structure of the isolated apical domain of *E. coli* GroEL has been determined to 1.7 Å resolution [12]. Like the complete protein, this domain binds and releases molten globules, but it lacks the cavities of the intact oligomer, so that the folding of released polypeptide chains takes place in free solution rather than inside an Anfinsen folding cage. By producing this fragment, Buckle *et al.* [12] have essentially reversed evolution by converting a large chaperone into a small chaperone; for this reason their use of the term 'mini-chaperone' is appropriate.

The apical domain was made with an additional amino-terminal tag of seventeen residues, six of which are histidines. Seven of the other residues — sequence GLVPRGS — were resolved in the structure and bind to nonpolar residues in a neighbouring domain in the crystal. Buckle *et al.* [12] suggest that this mimics the binding of a natural peptide substrate to an intact oligomer, but this conclusion awaits confirmation. The binding residues in the fragment are nonpolar, and are located in a shallow cleft between α helices H8 and H9 and in an adjacent surface formed by the packing between helix H9 and a neighbouring loop (Figure 1b). Most binding interactions are nonpolar, with four hydrogen bonds. Earlier mutational studies identified five of the binding residues as required for peptide binding, but also identified another four residues not seen in this study, so the binding site for natural peptide substrates probably extends further. The peptide binding site in GroEL is known to be flexible, so Buckle *et al.* [12] propose that it can accommodate a wide range of side chains as a result of movements of helices H8 and H9 and the nearby loops.

One difference between the peptide binding site of GroEL and that of DnaK is the absence of anything in the former corresponding to the latter's flexible lid. This difference can be rationalised on the basis that, in a cell, the apical domain of GroEL forms part of the inner surface of the folding cage, so peptides bound in this region are already sequestered from contact with peptides on other molten globules outside the cage. Once released from the apical domains, however, molten globules will diffuse rapidly into the cytoplasm from cavities that are uncapped by GroES. This effect has been observed *in vitro* [10], so how do cells circumvent this danger?

It is possible to mimic *in vitro* the crowded environment characteristic of cytoplasm by adding high concentrations of synthetic macromolecules such as dextran 70 and Ficoll [1]. The addition of such crowding agents to GroEL/ES-assisted protein folding mixtures almost completely abolishes the release of molten globules from uncapped cavities and their rebinding to other GroEL molecules [13]. The same result is observed if cytoplasmic extracts from *Xenopus* eggs are added, provided that the final concentration of protein is at least 200 mg ml^{-1} [13]. These observations suggest that, in the cell, a released molten globule rebinds to the same GroEL molecule that has just released it; this rapid recapture reduces the probability of aggregation with other molten globules. An additional effect of crowding is to reduce the diffusion of proteins by about one order of magnitude [1]. This reduction will also contribute to solving the essential problem that folding proteins face — how to avoid the crowded cytoplasm until their aggregation-prone surfaces are safely buried.

References

1. Zimmermann SB, Minton AP: **Macromolecular crowding: biochemical, biophysical and physiological consequences.** *Annu Rev Biomol Struct* 1993, **22:**27-65.
2. Zimmermann SB, Trach SO: **Estimation of the macromolecular concentrations and excluded volume effects for the cytoplasm of *Escherichia coli.*** *J Mol Biol* 1991, **222:**599-620.
3. Hartl FU: **Molecular chaperones in cellular protein folding.** *Nature* 1996, **381:**571-580.
4. Martin J, Hartl FU: **Chaperone-assisted protein folding.** *Curr Opin Struct Biol* 1997, **7:**41-52.
5. Ellis RJ: **Roles of molecular chaperones in protein folding.** *Curr Opin Struct Biol* 1994, **4:**117-122.
6. Horwich AL, Low KB, Fenton FA, Hirshfield IN, Furtak K: **Folding *in vivo* of bacterial cytoplasmic proteins: role of GroEL.** *Cell* 1993, **74:**909-917.
7. Ellis RJ (ed): *The Chaperonins.* San Diego: Academic Press; 1996:1-323.
8. Jackson GS, Staniforth RA, Halsall DJ, Atkinson T, Holbrook JJ, Clarke AR, Burston SG: **The binding and hydrolysis of nucleotides in the chaperonin catalytic cycle – implications for the mechanisn of assisted protein folding.** *Biochemistry* 1993, **32:**2554-2563.
9. Ellis RJ: **Opening and closing the Anfinsen cage.** *Curr Biol* 1994, **1:**633-635.
10. Weismann JS, Yashi Y, Fenton WA, Horwich AL: **GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms.** *Cell* 1994, **78:**693-702.
11. Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, Hendrickson WA: **Structural analysis of substrate binding by the molecular chaperone DnaK.** *Science* 1996, **272:**1606-1614.
12. Buckle AM, Zahn R, Fersht AR: **A structural model for GroEL–polypeptide recognition.** *Proc Natl Acad Sci USA* 1997, **94:**3571-3575.
13. Martin J, Hartl FU: **The effect of macromolecular crowding on chaperonin-mediated protein folding.** *Proc Natl Acad Sci USA* 1997, **94:**1107-1112.